Measurement of work in single-molecule pulling experiments

Alessandro Mossa,[1,*] Sara de Lorenzo,[1,2] Josep Maria Huguet,[1] and Felix Ritort[1,2,†]

[1] *Departament de Física Fonamental, Facultat de Física, Universitat de Barcelona, Avinguda Diagonal 647, 08028 Barcelona, España*

[2] *CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, Madrid, España*

A main goal of single-molecule experiments is to evaluate equilibrium free energy differences by applying fluctuation relations to repeated work measurements along irreversible processes. We quantify the error that is made in a free energy estimate by means of the Jarzynski equality when the accumulated work expended on the whole system (including the instrument) is erroneously replaced by the work transferred to the subsystem consisting of the sole molecular construct. We find that the error may be as large as 100%, depending on the number of experiments and on the bandwidth of the data acquisition apparatus. Our theoretical estimate is validated by numerical simulations and pulling experiments on DNA hairpins using optical tweezers.

PACS numbers: 05.70.Ln, 82.37.Rs, 87.80.Nj
Keywords: Jarzynski equality; single-molecule experiments; nonequilibrium thermodynamics

I. INTRODUCTION

In a typical single-molecule pulling experiment[1], an individual molecular construct is stretched by means of a device (e.g., optical or magnetic tweezers, atomic force microscope (AFM), etc.) able to measure both the applied force, usually on the piconewton scale, and the end-to-end molecular extension, typically expressed in nanometers. Many interesting kinetic and thermodynamical properties[2,3,4] of the stretching process can be inferred from the resulting force-extension curve (henceforth, FEC); in particular, the free energy difference between the folded and the unfolded state can be evaluated by exploiting a well-known result of nonequilibrium thermodynamics, the Jarzynski equality[5]:

$$\beta W_{\rm rev} = -\log\langle \exp[-\beta W(\Gamma)]\rangle_\Gamma \,, \qquad (1)$$

where $W(\Gamma)$ is the amount of work performed on the system throughout the stretching process Γ, β is as usual the inverse of the thermal energy $k_{\rm B}T$, and $W_{\rm rev}$ is the reversible work, i.e., the work needed to perform the pulling experiment in quasi-equilibrium conditions. Since a single molecule is a small system[6,7,8], $W(\Gamma)$ is affected by thermal fluctuations; the angular brackets $\langle\cdots\rangle_\Gamma$ thus stand for an average over all possible realizations of the same experimental protocol. In fact, a generalization of the Jarzynski equality due to Hummer and Szabo[9,10] makes it possible to reconstruct the whole free energy landscape as a function of the molecular extension[11,12,13]. This program has been successfully applied to the experimental study of multi-domain proteins[14,15].

Many a research has been devoted to the practical difficulties that arise when Eq. (1) is applied to the free energy reconstruction problem, e.g., the bias induced by the finite number of experimental attempts[16], the role played by the resolution of the measuring apparatus[17], or the effect of instrument noise and experimental errors[18]. The present article deals with yet another possible source of error, which, though already known, has generally been dismissed as negligible without a compelling argument. The point is that in most experimental settings the molecular extension is *not* the proper control parameter, so that it is not correct to interpret the area below the FEC as the work that appears in Eq. (1)[19]. If the control parameter is the total distance the area under the force-distance curve (FDC) should be used instead.

Here we thoroughly analyze under which conditions the use of the wrong definition for the work can appreciably affect the estimate of free energy differences by means of Eq. (1). The conclusion, in a nutshell, is that the error induced by the substitution may be as large as 100%, depending on the number of experiments and on the data acquisition frequency. Also important are the details of the data analysis procedure: how the integration extrema are chosen, what method is used to integrate the FEC and how different FECs are aligned to correct for instrumental drift effects.

The paper is organized as follows: First, we get some theoretical insight by considering our problem in its simplest possible setting (Sec. II). Then, we validate our conclusions with an experimental test implemented with optical tweezers and DNA hairpins (Sec. III). A recapitulation of our results (Sec. IV) and an appendix with some technicalities round off this article.

II. A TOY MODEL

A detailed model for single-molecule experiments with optical tweezers has been discussed elsewhere[3]. Here we consider a simplified version of it, that conserves only the physical features directly relevant to our problem. Although the toy model in this section is phrased in the optical tweezers language, it takes no effort to translate it into an AFM nomenclature, the mathematics being just the same.

In our model, graphically depicted in Fig. 1, the optical trap is moved by the experimenter, hence the proper control parameter is the trap–pipette distance λ, while

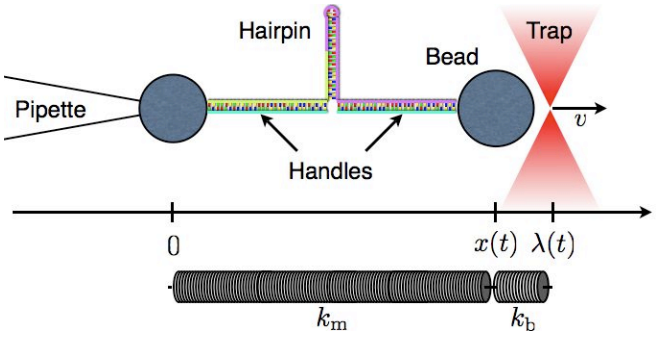

FIG. 1: Schematic definition of the model under study. The pipette is at rest with respect to the thermal bath, while the trap is moving with velocity v. The trap and the system molecule + handles are approximated by two harmonic potentials with stiffness k_b and k_m, respectively. The rest length of the trap spring k_b is zero, while the rest length of the molecule spring k_m is ℓ_0 if the hairpin is closed ($\varsigma = 0$) and ℓ_1 if it is open ($\varsigma = 1$).

the end-to-end molecular extension is a quantity subject to fluctuations denoted by x. The trap is an harmonic potential with stiffness k_b, while k_m is the stiffness of the molecular construct comprising hairpin and handles. Given a fixed value of the control parameter λ, the state of the system is specified by the pair (x, ς), where ς is a label taking values 0 if the hairpin is closed (or folded) and 1 if it is open (or unfolded). The hairpin itself is a pure two-state system[20] whose state-dependent length is ℓ_ς. The bead is thus subject to the net force

$$f_\mathrm{t}(x,\varsigma) = k_\mathrm{b}(\lambda - x) - k_\mathrm{m}(x - \ell_\varsigma)\,. \tag{2}$$

It is convenient to introduce the total stiffness $k_\mathrm{t} \equiv k_\mathrm{b} + k_\mathrm{m}$ and the equilibrium position (defined by the condition $f_\mathrm{t}(x_\mathrm{eq}, \varsigma) = 0$)

$$x_\mathrm{eq}(\varsigma) = \frac{k_\mathrm{b}\lambda + k_\mathrm{m}\ell_\varsigma}{k_\mathrm{t}}\,, \tag{3}$$

so that Eq. (2) can be rewritten as

$$f_\mathrm{t}(x,\varsigma) = -k_\mathrm{t}[x - x_\mathrm{eq}(\varsigma)]\,. \tag{4}$$

The relaxation time of the velocity autocorrelation function $\tau = m/\gamma$ (m being the mass and γ the friction coefficient of the bead in the trap) is small enough compared to the duration of the experiment that we can assume *mechanical equilibrium*[21], i.e. the average value of the total force $\langle f_\mathrm{t}(t)\rangle$ is zero. The Hamiltonian function is given by

$$H^{(\lambda)}(x,\varsigma) = \tfrac{1}{2}k_\mathrm{b}(\lambda - x)^2 + \tfrac{1}{2}k_\mathrm{m}(x - \ell_\varsigma)^2 + \varsigma\Delta G_0\,, \tag{5}$$

where ΔG_0 is the free energy difference between the open and closed states of the hairpin in the absence of applied force. The analytic solution to the equilibrium thermodynamics of this model is summarized in App. A.

The transitions of the hairpin are governed by a simplified Kramers–Bell kinetics[22], with rates for opening $k_{\rightarrow}$ or closing $k_{\leftarrow}$ given by

$$k_{\rightarrow} = k_0 \exp\left(\frac{w_0 f_0(x)}{k_\mathrm{B}T}\right)\,, \tag{6a}$$

$$k_{\leftarrow} = k_0 \exp\left(\frac{-w_1 f_1(x) + \Delta G_0}{k_\mathrm{B}T}\right)\,, \tag{6b}$$

where w_0 and w_1 represent the distances from the barrier to the closed and the open states, respectively, f_0 and f_1 are two functions of x with physical dimensions of a force, and k_0 is the attempt frequency. The rates just defined must respect the detailed balance condition

$$\frac{k_{\rightarrow}}{k_{\leftarrow}} = \exp\left[-\frac{H^{(\lambda)}(x,1) - H^{(\lambda)}(x,0)}{k_\mathrm{B}T}\right]\,, \tag{7}$$

for each λ and for each x. This requirement implies

$$w_0 f_0(x) + w_1 f_1(x) = \frac{k_\mathrm{m}}{2}(\ell_1 - \ell_0)[2x - (\ell_1 + \ell_0)]\,. \tag{8}$$

Our choice here is to take simply $f_0(x) = f_1(x)$, so that $w_0 + w_1 = \ell_1 - \ell_0$.

The dynamics of our model is ruled by the overdamped Langevin equation

$$\gamma\frac{\mathrm{d}x}{\mathrm{d}t} = f_\mathrm{t}(x(t),\varsigma) + \sqrt{2\gamma k_\mathrm{B}T}\xi(t)\,, \tag{9}$$

where $\xi(t)$ is a Gaussian white noise

$$\langle\xi(t)\rangle = 0\,, \tag{10a}$$

$$\langle\xi(t)\xi(t')\rangle = \delta(t - t')\,. \tag{10b}$$

The experimental protocol is defined by the choice of a function $\lambda(t)$. Here we consider a constant velocity pulling: $\lambda(t) = \lambda_0 + vt$.

A. Accumulated vs. transferred work

For the toy model introduced in the previous section, λ is the control parameter, which can be directly manipulated, while the molecular extension x is subject to Brownian fluctuations. Therefore, the work performed on the system throughout a pulling experiment Γ that starts at time t_i from $\lambda = \lambda_\mathrm{i}$ and terminates in $\lambda = \lambda_\mathrm{f}$ at time $t_\mathrm{f} = t_\mathrm{i} + (\lambda_\mathrm{f} - \lambda_\mathrm{i})/v$ is properly defined as

$$W(\Gamma) \equiv \int_{\lambda_\mathrm{i}}^{\lambda_\mathrm{f}} \frac{\partial H^{(\lambda)}(x,\varsigma)}{\partial\lambda}\,\mathrm{d}\lambda = \int_{\lambda_\mathrm{i}}^{\lambda_\mathrm{f}} f_\mathrm{b}(\lambda,x)\,\mathrm{d}\lambda\,. \tag{11}$$

where we used Eq. (5) and $f_\mathrm{b}(\lambda,x) \equiv k_\mathrm{b}(\lambda - x)$ is the force induced by the displacement of the bead in the trap. Such work is measured in practice as the area under the force-distance curve [FDC, see Fig. 2(a)]. Note that for all single-molecule techniques that we are aware of, f_b is actually the only one force experimentally measurable.

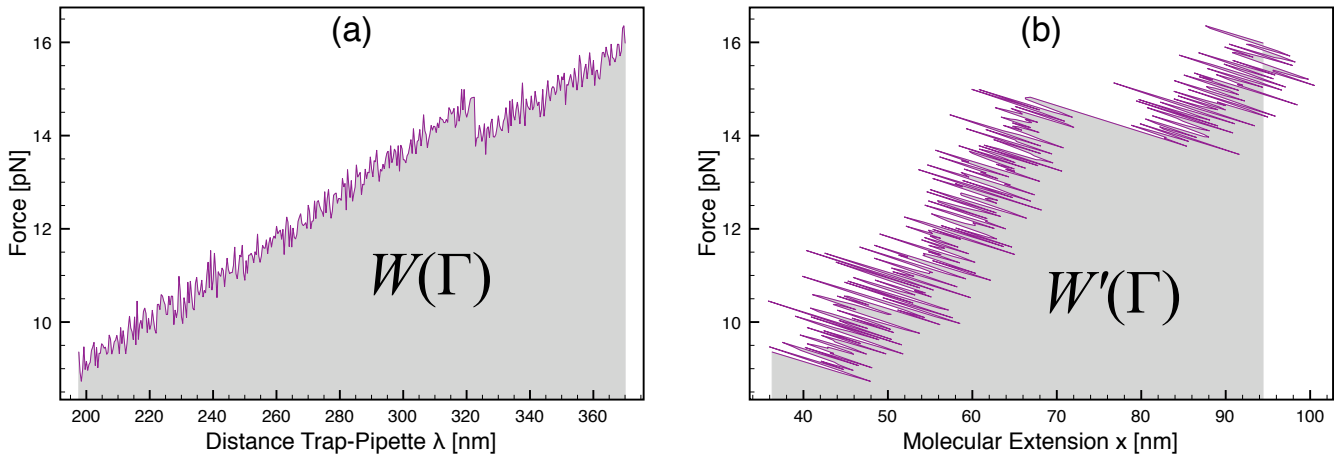

FIG. 2: (a) A typical force-distance curve (FDC) obtained by numerical simulation of Eq. (9). The shaded area is equivalent to the accumulated work $W(\Gamma)$ [see Eq. (11)]. (b) The force-extension curve (FEC) associated to the pulling experiment represented in Fig. 2(a). The shaded area is equivalent to the transferred work $W'(\Gamma)$ [see Eq. (12)].

In the following we will for simplicity drop the subscript and write f instead of f_b.

The area under the FEC [see Fig. 2(b)], on the other hand, is what in Ref. 19 is called transferred work [as opposed to the accumulated work $W(\Gamma)$]:

$$W'(\Gamma) \equiv \int_{x_\mathrm{i}(\Gamma)}^{x_\mathrm{f}(\Gamma)} f(\lambda, x)\, \mathrm{d}x\,, \tag{12}$$

where x_i and x_f are the *trajectory-dependent* values of the molecular extension at times t_i and t_f, respectively.

At each point along the trajectory Γ, the control parameter and the molecular extension are related by

$$x = \lambda - \frac{f}{k_\mathrm{b}}\,. \tag{13}$$

This implies the following relation between the area under a FDC and the area under the corresponding FEC:

$$W(\Gamma) = W'(\Gamma) + \frac{f_\mathrm{f}(\Gamma)^2 - f_\mathrm{i}(\Gamma)^2}{2k_\mathrm{b}}\,, \tag{14}$$

where f_i and f_f are the (trajectory-dependent) initial and final values of the force, respectively. The difference between W and W' is therefore a pure *boundary* term.

B. The reversible work

If we realize the pulling experiment in conditions of quasi-equilibrium, that is at infinitesimally small velocity $v \to 0$, then we obtain the thermodynamic force-distance curve (TFDC), whose analytical expression is given by Eq. (A5). The area under the TFDC is the reversible work W_rev, equal to the free energy difference between the final and initial states of the system. From an experimental perspective, however, the really interesting quantity is rather the free energy difference ΔG_0 between the open and closed states of the hairpin at zero external force. According to Eq. (A13), this is given by

$$\Delta G_0 = W_\mathrm{rev} - \frac{\langle f\rangle_\mathrm{f}^2 - \langle f\rangle_\mathrm{i}^2}{2k_\mathrm{eff}}\,, \tag{15}$$

where $\langle f\rangle_{\mathrm{i(f)}}$ is the equilibrium initial (final) value of the force, and k_eff is the effective stiffness

$$\frac{1}{k_\mathrm{eff}} = \frac{1}{k_\mathrm{b}} + \frac{1}{k_\mathrm{m}}\,. \tag{16}$$

The thermodynamic force-extension curve[3] (TFEC) is the quasi-equilibrium pulling experiment plotted as a function of the molecular extension x. If we define W'_rev as the area under the TFEC, then Eq. (14) yields

$$\Delta G_0 = W'_\mathrm{rev} - \frac{\langle f\rangle_\mathrm{f}^2 - \langle f\rangle_\mathrm{i}^2}{2k_\mathrm{m}}\,. \tag{17}$$

So we see that either W_rev or W'_rev are equally useful to extract the free energy of formation ΔG_0 of the hairpin. The problem is that it is often unpractical (and sometimes impossible) to achieve quasi-equilibrium conditions. Here comes into play the Jarzynski equality, as we see in the next section.

C. Jarzynski estimator

The Jarzynski equality Eq. (1) gives us a recipe to compute the reversible work, given a suitably-sized collection of irreversible processes. The work that appears in Eq. (1) is the accumulated work $W(\Gamma)$ defined in Eq. (11); nonetheless, in some cases it happens that the most readily available data for the experimenter is the FEC, therefore the work that is measured is in fact

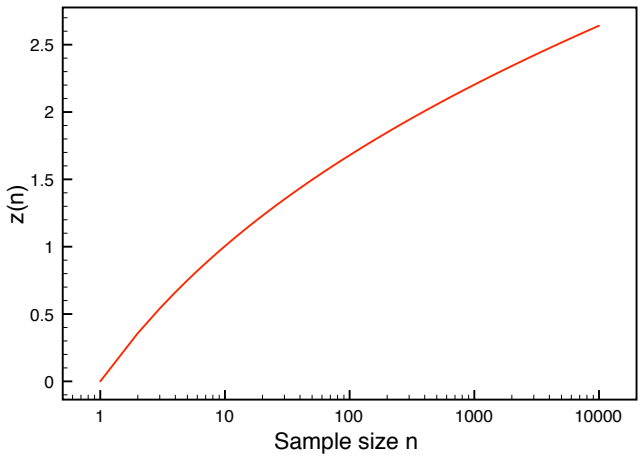

FIG. 3: Dependence on the sample size n of the mode ω' of $W'_{(1)}$ (i.e., the maximum of the distribution for $W'_{(1)}$, see App. B). The dimensionless variable z is $(\mu - \omega')/(\sqrt{2}\sigma)$, where μ and σ are the mean and standard deviation of the normally distributed transferred work w'. The represented curve is the numerical solution to Eq. (B9).

the transferred work $W'(\Gamma)$ of Eq. (12). In such occasions, the transferred work has been used in the Jarzynski equality, under the assumption that the resulting error is small compared to other sources of experimental uncertainty[23,24].

In this section, we answer the following question: How large an error in the evaluation of ΔG_0 is made if the transferred work $W'(\Gamma)$ is used instead of the accumulated work $W(\Gamma)$?

Let us call $\widetilde{W}$ the Jarzynski estimate of the reversible work $W_{\rm rev}$, based on n experiments that produce the set of work measurements $\{W_i\}$:

$$\beta\widetilde{W} \equiv -\log \sum_{i=1}^{n} \frac{1}{n} \exp(-\beta W_i)\,. \tag{18}$$

The analogous quantity obtained using the transferred work is

$$\beta\widetilde{W}' \equiv -\log \sum_{i=1}^{n} \frac{1}{n} \exp(-\beta W'_i)\,. \tag{19}$$

The quantity $\widetilde{W}$ is guaranteed by Eq. (1) to be an estimator of the reversible accumulated work $W_{\rm rev}$, whereas $\widetilde{W}'$ is *not* the proper way to compute the reversible transferred work $W'_{\rm rev}$ (a bona fide way to estimate $W'_{\rm rev}$ is discussed in Ref. 19). We now set out to evaluate the difference $\widetilde{W} - \widetilde{W}'$.

To begin with, we sort the set $\{W_i\}$ in ascending order:

$$W_{(1)} \le W_{(2)} \le W_{(3)} \le \cdots \le W_{(n)}\,. \tag{20}$$

The key observation is that the sum of exponentials in Eq. (18) is dominated by the minimum work trajectory of our sample:

$$\beta\widetilde{W} \approx \beta W_{(1)} + \log n\,. \tag{21}$$

Repeating the same argument for the set $\{W'_i\}$ that collects the measured values of the transferred work, we find

$$\widetilde{W} - \widetilde{W}' \approx W_{(1)} - W'_{(1)}\,. \tag{22}$$

Note that the trajectory that realizes the minimum of $\{W_i\}$ is generally not the same that gives the minimum of $\{W'_i\}$.

In order to go further in our analytical approximation, we need to specify the distributions of W and W'. Based on our experience with both experimental and simulated data, we assume that W' is normally distributed (see Fig. 8) with mean μ and variance σ^2, while for W we adopt a Gumbel distribution (see Fig. 9) with parameters a and b [which are related to the average and standard deviation of the accumulated work W by means of Eq. (B12) in App. B]. This latter choice is the simplest distribution that exhibits the asymmetry we expect from a nonlinear system[25] (in the case of linear systems the work distribution is Gaussian[26,27]). Also, there are theoretical arguments suggesting that the Gumbel distribution may play a universal role for correlated random variables similar to the one played by the Gaussian distribution for uncorrelated ones[28,29].

We can now estimate the distribution of $W_{(1)}$ and $W'_{(1)}$. The details can be found in App. B, here we quote just the final result: the most likely value of $W_{(1)} - W'_{(1)}$ is approximately

$$a - b\log n - \mu + \sqrt{2}\sigma z(n)\,, \tag{23}$$

where $z(n)$ is the function of the sample size represented in Fig. 3.

What we are really interested in, however, is ΔG_0. If we put $W_{\rm rev} = \widetilde{W}$ in Eq. (15) and call $\Delta G'_0$ the result of setting $W'_{\rm rev} = \widetilde{W}'$ in Eq. (17), we get

$$\Delta G_0 - \Delta G'_0 \approx a - b\log n - \mu + \sqrt{2}\sigma z(n) - \frac{\langle f\rangle_{\rm f}^2 - \langle f\rangle_{\rm i}^2}{2k_{\rm b}}\,. \tag{24}$$

A further simplification is possible: taking the average of Eq. (14) and using Eq. (B12) we are left with the formula

$$\Delta G_0 - \Delta G'_0 \approx \frac{\sqrt{6}}{\pi}(\gamma - \log n)s + \sqrt{2}z(n)s'\,, \tag{25}$$

where s and s' are the standard deviations of $\{W_i\}$ and $\{W'_i\}$, respectively, and γ is the Euler–Mascheroni constant.

Equation (25) states that the error in the evaluation of the energy properties of the hairpin due to the substitution of $\{W_i\}$ with $\{W'_i\}$ in the Jarzynski equation depends on three factors: the standard deviations s and s', and the number of experiments n. There is a remarkable difference between the roles played by s and s': the

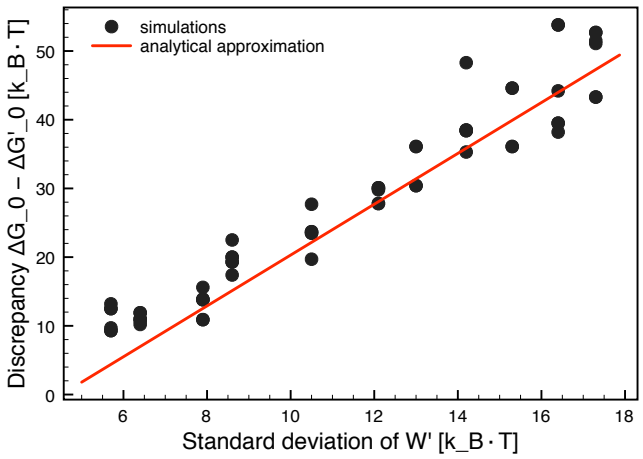

FIG. 4: Numerical test of Eq. (25). The theoretical prediction is compared to the results of numerical simulations of Eq. (9). In abscissa, s' is the standard deviation of the transferred work values $\{W'_i\}$; different values of s' are obtained by varying the filter applied to the data. In ordinate, the error $\Delta G_0 - \Delta G'_0$ (in $k_B T$ units) on the determination of the free energy of formation of the hairpin due to the erroneous use of W' in the Jarzynski estimator. Each point represents the result of the analysis of $n = 9000$ trajectories.

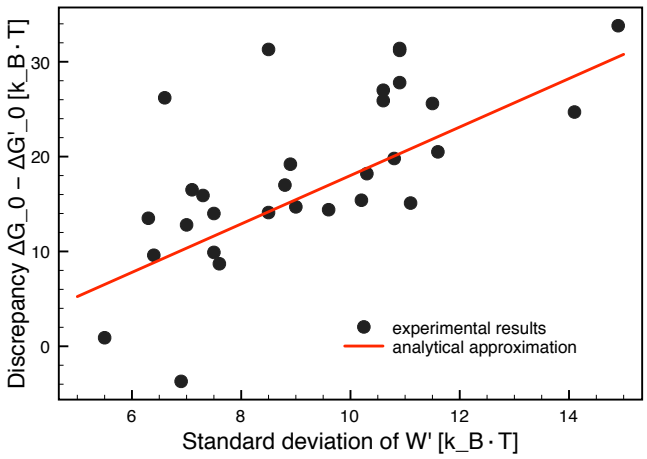

FIG. 5: Experimental test of Eq. (25). In abscissa, s' is the standard deviation of the transferred work values $\{W'_i\}$; different values of s' are obtained by varying the stiffness of the trap and the bandwidth. In ordinate, the error $\Delta G_0 - \Delta G'_0$ on the determination of the hairpin energy levels due to the erroneous use of W' in the Jarzynski estimator. See Tab. I for further details about the data.

standard deviation s of the accumulated work generally depends only on the pulling rate v and the chemical nature of the construct comprising molecule and handles; the standard deviation s' of the transferred work, on the other hand, is also strongly dependent on the bandwidth of the data acquisition system.

The reason is easy to understand: while the area under the FDC [Fig. 2(a)] practically doesn't change if we smooth out the curve, the area under the FEC [Fig. 2(b)] is heavily dependent on the fluctuations of the extremal points x_i and x_f (see also Fig. 6). We will have more to say about this point in Sec. III.

In the derivation of Eq. (25) we have made use of three approximations:

- we discarded all the contributions to the sum of exponentials in Eqs. (18) and (19) except the one coming from the minimum-work trajectory;
- we assumed a normal distribution for $\{W'_i\}$;
- we assumed a Gumbel distribution for $\{W_i\}$.

Although each one of them seems reasonable, it is not redundant, before discussing the experimental utility of Eq. (25), to check the final result against a numerical test.

D. A numerical test

In order to validate Eq. (25), we have performed a numerical simulation of Eq. (9), generating hundreds of thousands of curves like the two represented in Fig. 2. The effect of the instrumental bandwidth has been mimicked by applying different filters to the data, so that each point of the FDC or FEC represents actually an average over m consecutive integration steps. In this way we have generated data in a fair range of values of s'. The results are illustrated in Fig. 4.

The first observation is that the error can be very large: as much as 50 $k_B T$ in a system where the true ΔG_0 is 57.7 $k_B T$, that amounts to a relative error not far from 100%. Then we observe that, in spite of the somewhat rough simplifications used in its derivation, the analytical prediction of Eq. (25) fares reasonably well in the comparison with the simulated data, although there seems to be a small apparently systematic underestimation of $\Delta G_0 - \Delta G'_0$. Finally, a comment about the range of s': The standard deviation of $\{W'_i\}$ is a linear function of the amplitude of the fluctuations of x, given by Eq. (A16); this fixes an upper limit to the range of s' that can be explored without changing the system.

III. AN EXPERIMENTAL TEST

This section reports the results of an experimental test of Eq. (25), whose theoretical derivation has been presented in Sec. II. The instrument we employed is a dual-beam miniaturized optical tweezers with fiber-coupled diode lasers (845 nm wavelength) that produce a piezo controlled movable optical trap and measure force using conservation of light momentum[30,31]. The molecule is a DNA hairpin of sequence 5'-GCGAGCCATAATCTC-ATCTGGAAACAGATGAGATTATGGCTCGC-3' hybridized to two double-stranded DNA (dsDNA) handles

(29 base-pairs long). Pulling experiments were performed at 25 °C in a buffer containing Tris H-Cl pH 7.5, 1 M EDTA and 1 M NaCl. The data that we show (see Tab. I) have been measured from 7 specimens in hundreds of stretching-releasing cycles performed at pulling speed of 200 nm/s (equivalent to a loading rate of 13.8 pN/s). The use of DNA hairpins presents several advantages[32,33,34,35] over the RNA hairpins that were used in pioneering experiments of this kind[23,24].

In order to measure the dependence of $\Delta G_0 - \Delta G'_0$ on the bandwidth, we employed a fast analog-to-digital converter that makes possible to increase the data acquisition frequency from the standard value of 1 kHz to as much as 100 kHz (20 kHz, however, is larger than the corner frequency of the bead, around 10 kHz, and proved to be enough for this test). The availability of high-frequency data is a good start, but is not enough without a data analysis procedure that carefully preserves the statistical properties of the boundary term [see Eq. (14)]. Here are the main steps of the data analysis that we performed:

1. The stream of data is split into single unfolding or refolding events.

2. Taking advantage of the fact that the elastic response of the short dsDNA handles is with a good approximation Hookean, we fit the FDC folded and unfolded branches with straight lines.

3. The unavoidable small instrumental drift (which is manifested in the unphysical increasing or decreasing of the measured value of the trap positon λ) is corrected by shifting the FDC in such a way that the straight line fitting the folded branch crosses $\lambda = 0$ at the same value of the force in any event.

4. The FDCs are integrated between two fixed values $\lambda_{\rm i}$ and $\lambda_{\rm f}$. These integrations produce two sets of accumulated work values $\{W_i\}$: one for the unfolding and one for the refolding process.

5. Each FEC is integrated between $x_{\rm i}(\Gamma) \equiv \lambda_{\rm i} - f_{\rm i}(\Gamma)/k_{\rm b}$ and $x_{\rm f}(\Gamma) \equiv \lambda_{\rm f} - f_{\rm f}(\Gamma)/k_{\rm b}$; note that, while $\lambda_{\rm i}$ and $\lambda_{\rm f}$ are the same for all trajectories, $f_{\rm i}$ and $f_{\rm f}$ depend on the trajectory Γ, and so do $x_{\rm i}$ and $x_{\rm f}$. In this way we obtain two sets of transferred work values $\{W'_i\}$: again, one for unfolding and one for refolding trajectories.

6. The Jarzynski estimators $\widetilde{W}$ and $\widetilde{W'}$ are computed by means of Eqs. (18) and (19), and then Eqs. (15) and (17) give ΔG_0 and $\Delta G'_0$.

Table I shows that Eq. (25) is generally quite close to the experimental results, most of the times predicting a discrepancy between ΔG_0 and $\Delta G'_0$ within few $k_{\rm B}T$ of the observed value. The occasional large deviations between theory and experiment shouldn't be too surprising in view of the statistical nature of the quantity we are measuring and the approximate derivation of Eq. (25).

The data reported in Tab. I can be graphically represented in analogy with Fig. 4. In principle, we expect each dataset to be represented by a slightly different straight line, as the number of trajectories n varies from a minimum 143 to a maximum 635 (see Tab. I). However, in practice the differences are small enough that all the theoretically expected values are very close to the line that in Fig. 5 is denoted as "analytical approximation".

Figure 6 shows a typical trajectory plotted as FDC and FEC, using 20 kHz and 1 kHz data. It can be immediately appreciated that, while the area under the FDC is insensitive to the sampling frequency, the area under the FEC may display important differences due to the fluctuations of the integration extrema.

A. Bi-directional methods

If the experimental situation makes it possible to implement not only the protocol $\lambda(t)$, but also the time-reversed protocol $\widehat{\lambda}(t) \equiv \lambda(\Delta t - t)$, where $\Delta t \equiv t_{\rm f} - t_{\rm i}$ is the duration of the experiment, then a more efficient way of estimating free energy differences is to apply a bi-directional method[24,36,37], which takes advantage of the knowledge of both a "forward" and a "reverse" work distributions. Bi-directional methods are based on another fluctuation relation, the Crooks theorem[38]

$$\frac{\phi_{W_{\rm FOR}}(w)}{\phi_{W_{\rm REV}}(-w)} = \exp\left(\frac{w - \Delta G}{k_{\rm B}T}\right) , \qquad (26)$$

where $\phi_{W_{\rm FOR}}(w)$ ($\phi_{W_{\rm REV}}(w)$) is the probability density function of the work along the forward (reverse) process. Also the Crooks theorem, like the Jarzynski equality, is written for the accumulated work W. Writing an analytical approximation of the error introduced by the erroneous use of the transferred work W', in the style of what we did in Sec. II, looks quite more complicated, but a direct evidence of the role of the bandwidth is given in Fig. 7, where $\log[\phi_{W_{\rm FOR}}(w)/\phi_{W_{\rm REV}}(-w)]$ is plotted as a function of $(w - \Delta G)/(k_{\rm B}T)$ for two values of the bandwith.

The experimental results are summarized in Tab. II. Even if the Crooks theorem is not satisfied, the estimate of ΔG_0 that we get by blindly substituting in Eq. (26) the transferred work W' for the accumulated work W is not as bad as the one obtained by using the Jarzynski equality.

B. Role of the data analysis technique

The data analysis protocol detailed in Sec. III may be the best suited to the task of verifying Eq. (25), but is not feasible if one's experimental setting only provides access to the transferred work W' (and makes it difficult to accurately estimate the stiffness $k_{\rm b}$ of the trap). If this is the case, then one either employs a version of

TABLE I: Experimental results: Comparison between the experimental (also shown in Fig. 5) and the theoretical (based on Eq. (25)) values of $(\Delta G_0 - \Delta G_0')/(k_B T)$. The datasets labeled "1 kHz" and "20 kHz" refer to the same experiment, with the standard (low-frequency) and the new (high-frequency) data acquisition system. The stiffness of the trap k_b is measured in pN/µm, while n is the number of trajectories.

		from unfolding			from refolding	
	k_b	n	exp.	th.	n	exp.	th.
mol1 1 kHz	79	249	16.5	11.6	249	8.7	12.2
mol1 20 kHz	79	200	31.4	20.6	202	15.1	20.2
mol2 1 kHz	63	473	25.6	23.3	473	20.6	22.8
mol2 20 kHz	63	364	24.7	29.5	362	33.8	31.2
mol3 1 kHz	79	219	15.9	17.5	218	12.8	14.7
mol3 20 kHz	79	169	31.2	20.1	166	19.8	19.0
mol4 1 kHz	64	174	31.3	13.8	174	−3.7	9.6
mol4 20 kHz	64	143	18.2	17.4	138	14.4	15.8
mol5 1 kHz	79	635	26.2	9.1	633	13.5	8.1
mol5 20 kHz	79	501	19.2	15.2	499	17.0	14.6
mol6 1 kHz	83	490	14.7	15.5	492	14.1	12.6
mol6 20 kHz	83	386	27.0	18.6	384	15.4	16.6
mol7 1 kHz	77	272	14.0	11.9	277	9.9	11.6
mol7 20 kHz	77	215	27.8	20.1	219	25.9	19.1

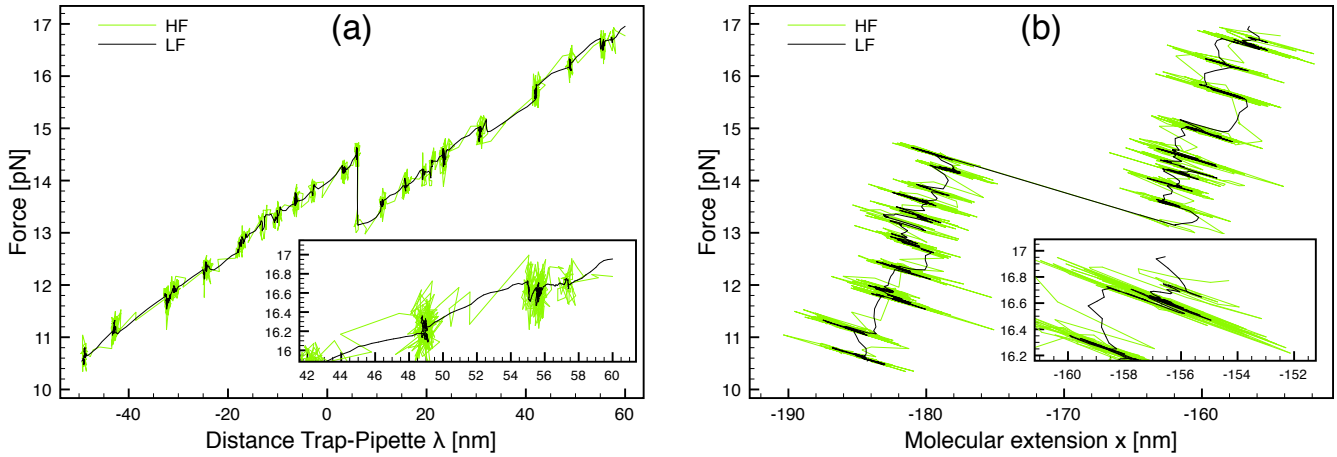

FIG. 6: (a) An experimental force-distance curve (FDC) observed with a high-frequency (20 kHz) and a low-frequency (1 kHz) data acquisition system. The area under the curve, which is a measure of the accumulated work W, practically doesn't change. (b) The force-extension curve (FEC) associated to the pulling experiment represented in Fig. 6(a). The area under the curve, which represents the transferred work W', depends on the frequency of the data acquisition system because of the large fluctuations of the integration extrema. Insets: magnified views of the region around the maximum of the force.

TABLE II: Experimental results. The datasets labeled "1 kHz" and "20 kHz" refer to the same experiment, with the standard (low-frequency) and the new (high-frequency) data acquisition system. The datasets labeled "ave n" are obtained from 20 kHz data by averaging over n points.

	from unfolding		from refolding		bi-directional	
	ΔG_0	$\Delta G_0'$	ΔG_0	$\Delta G_0'$	ΔG_0	$\Delta G_0'$
20 kHz	61.9	37.2	61.1	94.9	61.8	54.9
ave 2	62.0	40.2	61.2	92.6	61.6	55.3
ave 3	62.0	39.1	61.2	86.2	61.7	55.7
ave 4	61.7	41.2	60.9	86.3	61.3	55.5
ave 5	61.9	39.2	61.1	80.2	61.5	55.6
ave 10	61.8	42.4	61.1	76.2	61.4	56.1
ave 15	61.5	42.0	60.9	77.4	61.2	56.0
ave 20	61.6	41.5	60.9	77.2	61.2	56.2

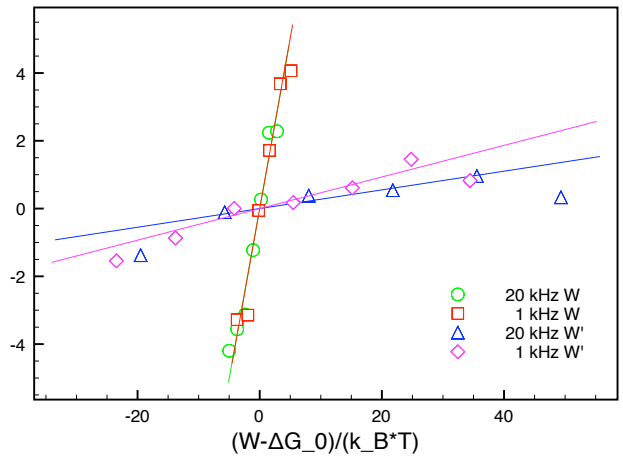

FIG. 7: Graph of $\log(\phi_{W_{\rm FOR}}(w)/\phi_{W_{\rm REV}}(-w))$ using high- and low-frequency data, accumulated and transferred work. Data have been shifted along the horizontal axis to be easily compared. Data for the accumulated work (circles and squares) fall into a (bandwidth-independent) straight line of slope 1.00(8) in quantitative agreement with the prediction by the fluctuation relation Eq. (26). However data for the transferred work (triangles and rhombs) exhibit bandwidth-dependent very small slopes (around 0.03) that exclude the validity of an equivalent relation to Eq. (26) for the transferred work.

the fluctuation theorem written for W' (as in the already cited Ref. 19), or uses W' in Eq. (1), but takes care of minimizing the error on the determination of ΔG_0, approximately given by Eq. (25). For example, the folded and unfolded branches of the FEC can be smoothed (by application of a filter, by spline-fitting, etc.) until the variance of $\{W_i'\}$ is entirely due to the distribution of the breaking point, in which case the difference between ΔG_0 and $\Delta G_0'$ becomes negligible compared to other sources of experimental error. This is the reason why both Refs. 23 and 24 obtained an acceptable experimental test of the Jarzynski equality and the Crooks theorem, respectively, even if erroneously using the transferred work.

IV. CONCLUSION

The output of a single-molecule pulling experiment can be graphically represented in the form of a force-extension curve, where the externally applied force is compared to the molecular construct end-to-end distance, or a force-distance curve, where the same force is represented against the physical control parameter, the length that can be directly manipulated by the experimenter. The area under the former curve is the work W' transferred to the molecule subsystem, while the latter curve allows the measurement of the accumulated work W, the total amount of work expended on the whole system (experimental apparatus included).

The fluctuation theorems commonly used to compute free energy differences from these out-of-equilibrium processes apply to the work W, but not to the work W'. In this paper we quantified how large an error is likely to affect the estimate of the free energy at zero force ΔG_0 of the molecule if W is erroneously replaced with W'. We found an analytical approximated expression [Eq. (25)] that emphasizes the role of the data analysis procedure and of the bandwidth of the data acquisition system. We confirmed the validity of this approach by both numerical simulation of a toy model and experiments on a DNA hairpin. This work should resolve some issues about the proper way to measure work in single-molecule experiments that have generated discussion and controversy over the past years.

Acknowledgments

The authors gratefully acknowledge financial support from grants FIS2007-61433, NAN2004-9348 from Spanish Research Council, SGR05-00688 from the Catalan Government and RGP55/2008 from Human Frontiers Science Program.

APPENDIX A: THERMODYNAMICS OF THE TOY MODEL

The model defined in Sec. II is simple enough to allow the analytical solution of its equilibrium thermodynamics. The partition function of the system is

$$Z(\lambda) = \sum_{\varsigma\in\{0,1\}} \int_{-\infty}^{+\infty} {\rm d}x \, \exp\left[-\beta H^{(\lambda)}(x,\varsigma)\right] , \qquad \text{(A1)}$$

where the Hamiltonian is given by Eq. (5). The integration is trivial, so we can immediately write the solution

$$Z(\lambda) = Z_0(\lambda) + Z_1(\lambda) , \qquad \text{(A2)}$$

where

$$Z_\varsigma(\lambda) = \sqrt{\frac{2\pi}{\beta k_{\rm t}}} \exp\left[-\frac{\beta}{2} k_{\rm eff}(\lambda-\ell_\varsigma)^2 - \varsigma\beta\Delta G_0\right] . \qquad \text{(A3)}$$

Given the partition function, we have access to all the thermodynamic properties of the model; the Gibbs free energy, in particular, is defined as

$$G(\lambda) = -k_{\rm B}T \ln Z(\lambda) , \qquad \text{(A4)}$$

and the TFDC is given by

$$\langle f\rangle(\lambda) = \frac{\partial G(\lambda)}{\partial\lambda} = k_{\rm eff}\left[\lambda - P_1(\lambda)\ell_1 - P_0(\lambda)\ell_0\right] , \qquad \text{(A5)}$$

where

$$P_\varsigma(\lambda) = \frac{Z_\varsigma(\lambda)}{Z(\lambda)} \qquad \text{(A6)}$$

is the probability of the state ς for a fixed value of λ. The coexistence value λ_c is characterized by the fact that $P_0(\lambda_c) = P_1(\lambda_c)$, hence

$$\lambda_c = \frac{\Delta G_0}{k_{\rm eff}(\ell_1 - \ell_0)} + \frac{\ell_1 + \ell_0}{2}\,. \tag{A7}$$

The corresponding coexistence force is

$$f_c \equiv \langle f\rangle(\lambda_c) = \frac{\Delta G_0}{\ell_1 - \ell_0}\,. \tag{A8}$$

Notice that in the asymptotic region $\lambda \ll \lambda_c$ the probability of the open state is negligible, so the force goes as $k_{\rm eff}(\lambda - \ell_0)$, while in the region $\lambda \gg \lambda_c$ it is the probability of the closed state that goes to zero, leaving a force dependence of the form $k_{\rm eff}(\lambda - \ell_1)$.

From Eq. (A5) we can easily write down the reversible work

$$W_{\rm rev} = \int_{\lambda_i}^{\lambda_f} \langle f\rangle(\lambda)\,{\rm d}\lambda\,. \tag{A9}$$

The integration can be done analytically using the fact that

$$\int \frac{a\,{\rm d}x}{a + {\rm e}^{bx}} = x - \frac{1}{b}\ln(a + {\rm e}^{bx})\,. \tag{A10}$$

Some tedious algebraic manipulation is required before one can write for the reversible work the following exact formula:

$$W_{\rm rev} = \Delta G_0 + \frac{k_{\rm eff}}{2}\left[(\lambda_{\rm f} - \ell_1)^2 - (\lambda_{\rm i} - \ell_0)^2\right] - C\,, \tag{A11}$$

where C is a correction very small if $\lambda_{\rm i} \ll \lambda_c \ll \lambda_{\rm f}$ (that is the most common experimental condition) whose explicit form is

$$C = \frac{1}{\beta}\ln\frac{1 + \exp[-\beta k_{\rm eff}(\ell_1 - \ell_0)(\lambda_{\rm f} - \lambda_c)]}{1 + \exp[-\beta k_{\rm eff}(\ell_1 - \ell_0)(\lambda_c - \lambda_{\rm i})]}\,. \tag{A12}$$

In practice, $\lambda_{\rm i} \ll \lambda_c \ll \lambda_{\rm f}$ so one can usually forget about C and use Eq. (A5) to rewrite Eq. (A11) as

$$W_{\rm rev} = \Delta G_0 + \frac{\langle f\rangle_{\rm f}^2 - \langle f\rangle_{\rm i}^2}{2k_{\rm eff}}\,, \tag{A13}$$

where $\langle f\rangle_{\rm i} \equiv \langle f\rangle(\lambda_{\rm i})$ and $\langle f\rangle_{\rm f} \equiv \langle f\rangle(\lambda_{\rm f})$.

The expectation value of the molecular extension is

$$\langle x\rangle(\lambda) = \frac{k_{\rm b}}{k_{\rm t}}\lambda + \frac{k_{\rm m}}{k_{\rm t}}\left[P_1(\lambda)\ell_1 + P_0(\lambda)\ell_0\right]\,. \tag{A14}$$

This equation can be rephrased into an expression for the TFEC.

Another interesting quantity is the expectation value of x^2,

$$\langle x^2\rangle(\lambda) = \frac{1}{\beta k_{\rm t}} + P_0 x_{\rm eq}^2(0) + P_1 x_{\rm eq}^2(1)\,, \tag{A15}$$

from which we easily obtain the variance for the equilibrium fluctuations of x

$$(\delta x)^2(\lambda) = \frac{k_{\rm B}T}{k_{\rm t}} + P_0(\lambda)P_1(\lambda)\frac{k_{\rm m}^2}{k_{\rm t}^2}(\ell_1 - \ell_0)^2\,. \tag{A16}$$

The variance for the equilibrium fluctuations of the force are simply related to those of x:

$$(\delta f)^2(\lambda) = k_{\rm b}^2(\delta x)^2(\lambda)\,. \tag{A17}$$

APPENDIX B: AN EXERCISE IN ORDER STATISTICS

Let $\{Y_i\}$ be n independent, identically distributed real-valued random variables with cumulative density function (cdf) $\Phi(y) \equiv \Pr(Y_i \le y)$. The probability density function (pdf) is defined as the derivative of the cdf: $\phi(y) \equiv \Phi'(y)$. The pdf has the property $\phi(y){\rm d}y = \Pr(y < Y_i \le y + {\rm d}y)$.

The minimum $Y_{(1)}$ of the set $\{Y_i\}$ is itself a random variable whose distribution can be deduced from the knowledge of $\phi(y)$ and $\Phi(y)$. Indeed, the probability $\Phi_{Y_{(1)}}(y)$ that the minimum is no more than y is equal to the probability of having at least one $Y_i \le y$. This is given by the binomial distribution as

$$\Phi_{Y_{(1)}}(y) = 1 - [1 - \Phi(y)]^n\,. \tag{B1}$$

Differentiating with respect to y we find the corresponding pdf

$$\phi_{Y_{(1)}}(y) = n[1 - \Phi(y)]^{n-1}\phi(y)\,. \tag{B2}$$

The simplest way to characterize the most likely value of $Y_{(1)}$ is to consider the mode, that is the point where the pdf has a maximum. This is given by solving with respect to y the following equation:

$$[1 - \Phi(y)]\phi'(y) = (n-1)\phi^2(y)\,. \tag{B3}$$

In the rest of this section, we specialize these general formulas to the two distributions we used to describe the statistical behavior of the accumulated and transferred work.

1. Normal distribution

A normally distributed variable of mean μ and variance σ^2 is described by the cdf

$$\Phi^{(\rm N)}(y) = \frac{1}{2} + \frac{1}{2}\,{\rm erf}\left(\frac{y - \mu}{\sqrt{2}\sigma}\right)\,, \tag{B4}$$

from which derives the pdf

$$\phi^{(\rm N)}(y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]\,. \tag{B5}$$

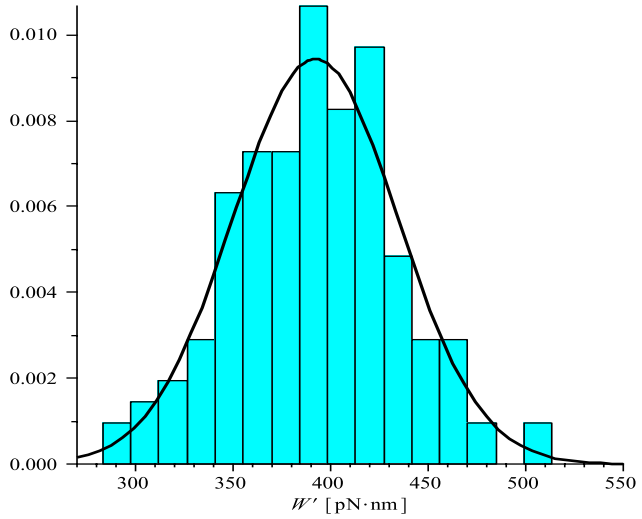

FIG. 8: Comparison between the histogram of the transferred work in one of the experiments reported in Tab. I and the normal distribution that better approximates it.

The distribution of the transferred work W' is often well described by a normal distribution (see Fig. 8). It is convenient to define the reduced variable

$$z \equiv \frac{\mu - y}{\sqrt{2}\sigma}\,, \tag{B6}$$

in terms of which we can write the cdf of the minimum $Y_{(1)}$ of a sample of size n

$$\Phi^{(\mathrm{N})}_{Y_{(1)}}(z) = 1 - \left[\tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}(z)\right]^n\,, \tag{B7}$$

and its pdf

$$\phi^{(\mathrm{N})}_{Y_{(1)}}(z) = \frac{n}{\sigma\sqrt{2\pi}}\exp(-z^2)\left[\tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}(z)\right]^{n-1}\,. \tag{B8}$$

The mode of the distribution $\phi^{(\mathrm{N})}_{Y_{(1)}}(z)$ is the solution to the following transcendental equation:

$$\sqrt{\pi}z[1 + \mathrm{erf}(z)] = (n-1)\exp(-z^2)\,. \tag{B9}$$

The numerical solution for $n \leq 10\,000$ is plotted in Fig. 3.

2. Gumbel distribution

In both our simulations and experiments, we find that the accumulated work is often adequately represented (see Fig. 9) by a random variable obeying the Gumbel distribution

$$\Phi^{(\mathrm{G})}(y) = 1 - \exp\left[-\exp\left(\frac{y-a}{b}\right)\right]\,, \tag{B10}$$

$$\phi^{(\mathrm{G})}(y) = \frac{1}{b}\exp\left(\frac{y-a}{b}\right)\exp\left[-\exp\left(\frac{y-a}{b}\right)\right]\,. \tag{B11}$$

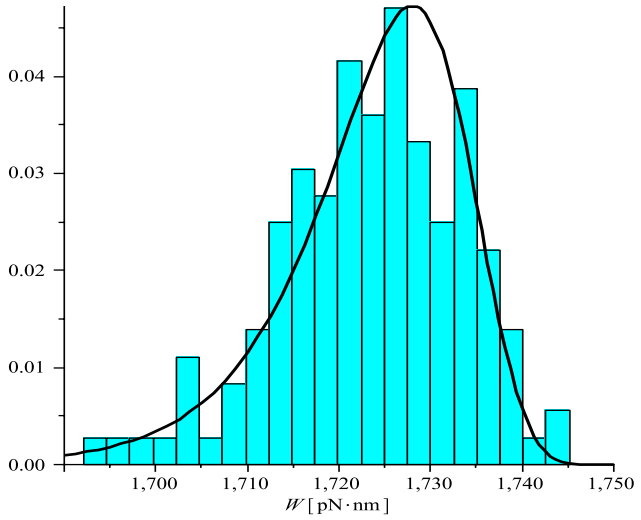

FIG. 9: Comparison between the histogram of the accumulated work in one of the experiments reported in Tab. I and the Gumbel distribution that better approximates it.

The parameters a and b can be quickly estimated from the average $\bar{y}$ and the standard deviation s of the sample $\{Y_i\}$ by means of the formulas

$$b = s\frac{\sqrt{6}}{\pi} \qquad a = \bar{y} + \gamma b\,, \tag{B12}$$

where γ is the Euler–Mascheroni constant 0.5772... The minimum value $Y_{(1)}$ over the sample is in this case distributed with pdf

$$\phi^{(\mathrm{G})}_{Y_{(1)}}(y) = \frac{n}{b}\exp\left(\frac{y-a}{b}\right)\exp\left[-n\exp\left(\frac{y-a}{b}\right)\right]\,. \tag{B13}$$

The mode of the minimum is therefore given simply by $a - b\log n$.

* Electronic address: mossa@ub.edu
† Electronic address: ritort@ffn.ub.es

[1] F. Ritort, J. Phys.: Condens. Matter **18**, R531 (2006), arXiv:cond-mat/0609378.
[2] C. Hyeon and D. Thirumalai, J. Phys.: Condens. Matter **19**, 113101 (2007), arXiv:cond-mat/0612433.

[3] M. Manosas and F. Ritort, Biophys. J. **88**, 3224 (2005), arXiv:cond-mat/0405035.
[4] M. Manosas, D. Collin, and F. Ritort, Phys. Rev. Lett. **96**, 218301 (2006), arXiv:cond-mat/0606254.
[5] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997), cond-mat/9610209.
[6] F. Ritort, Adv. Chem. Phys. **137**, 31 (2008), arXiv:0705.0455.
[7] U. Marini Bettolo Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani, Phys. Rep. **461**, 111 (2008), arXiv:0803.0719.
[8] F. Ritort, C. R. Physique **8**, 528 (2007).
[9] G. Hummer and A. Szabo, Proc. Nat. Acad. Sci. USA **98**, 3658 (2001).
[10] G. Hummer and A. Szabo, Acc. Chem. Res. **38**, 504 (2005).
[11] O. Braun, A. Hanke, and U. Seifert, Phys. Rev. Lett. **93**, 158105 (2004), arXiv:cond-mat/0402496.
[12] A. Imparato and L. Peliti, J. Stat. Mech. (2006) P03005, arXiv:cond-mat/0601552.
[13] C. Hyeon, G. Morrison, and D. Thirumalai, Proc. Nat. Acad. Sci. USA **105**, 9604 (2008), arXiv:0808.0480.
[14] N. C. Harris, Y. Song, and C.-H. Kiang, Phys. Rev. Lett. **99**, 068101 (2007), arXiv:0707.0504.
[15] A. Imparato, F. Sbrana, and M. Vassalli, Europhys. Lett. **82**, 58006 (2008), arXiv:0804.2980.
[16] J. Gore, F. Ritort, and C. Bustamante, Proc. Nat. Acad. Sci. USA **100**, 12564 (2003).
[17] S. Rahav and C. Jarzynski, J. Stat. Mech. (2007) P09012, arXiv:0708.2437.
[18] P. Maragakis, F. Ritort, C. Bustamante, M. Karplus, and G. E. Crooks, J. Chem. Phys. **129**, 024102 (2008), arXiv:0707.0089.
[19] J. M. Schurr and B. S. Fujimoto, J. Phys. Chem. B **107**, 14007 (2003).
[20] F. Ritort, C. Bustamante, and I. Tinoco, Jr., Proc. Nat. Acad. Sci. USA **99**, 13544 (2002), arXiv:physics/0210063.
[21] R. D. Astumian, J. Chem. Phys. **126**, 111102 (2007), arXiv:cond-mat/0608352.
[22] I. Tinoco, Jr., Annu. Rev. Biophys. Biomol. Struct. **33**, 363 (2004).
[23] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Science **296**, 1832 (2002).
[24] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Nature **437**, 231 (2005), arXiv:cond-mat/0512266.
[25] A. Saha and J. K. Bhattacharjee, J. Phys. A: Math. Theor. **40**, 13269 (2007), arXiv:cond-mat/0612423.
[26] F. Douarche, S. Ciliberto, A. Petrosyan, and I. Rabbiosi, Europhys. Lett. **70**, 593 (2005), arXiv:cond-mat/0502395.
[27] F. Douarche, S. Ciliberto, and A. Petrosyan, J. Stat. Mech. (2005) P09011, arXiv:cond-mat/0504465.
[28] E. Bertin, Phys. Rev. Lett. **95**, 170601 (2005), arXiv:cond-mat/0506166.
[29] E. Bertin and M. Clusel, J. Phys. A: Math. Gen. **39**, 7607 (2006), arXiv:cond-mat/0601189.
[30] C. Bustamante and S. B. Smith, U.S. Patent 7 133 132, B2 (2006).
[31] S. B. Smith, Y. Cui, and C. Bustamante, in *Biophotonics, Part B*, edited by G. Marriott and I. Parker (Academic Press, 2003), vol. 361 of *Methods in Enzymology*, pp. 134–162.
[32] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, Proc. Nat. Acad. Sci. USA **103**, 6190 (2006).
[33] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block, Science **314**, 1001 (2006).
[34] A. Mossa, M. Manosas, N. Forns, J. M. Huguet, and F. Ritort, J. Stat. Mech. (2009) P02060, arXiv:0902.3632.
[35] M. Manosas, A. Mossa, N. Forns, J. M. Huguet, and F. Ritort, J. Stat. Mech. (2009) P02061, arXiv:0902.3634.
[36] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).
[37] D. D. L. Minh and A. B. Adib, Phys. Rev. Lett. **100**, 180602 (4 pages) (2008), arXiv:0802.0224.
[38] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999), arXiv:cond-mat/9901352.